4

ANOVA and ANCOVA: Field Competition Experiments

Deborah E. Goldberg and Samuel M. Scheiner

4.1 Introduction

Competition has occupied a central place in ecological theory since Darwin and experiments on competition have been conducted on many different organisms from many different environments (see reviews in Jackson, 1981; Connell, 1983; Schoener, 1983; Hairston, 1989; Goldberg and Barton, 1992). There are many different kinds of competition experiments. The main focus of this chapter is how to choose the appropriate designs and statistical analyses for addressing particular kinds of questions about competition. Such choices depend on many aspects of the questions and systems under study. Because the basic statistical methods appropriate for most of the designs we present, analysis of variance (ANOVA) and analysis of covariance (ANCOVA), are thoroughly covered in standard textbooks of experimental design and analysis, we provide somewhat less emphasis on the nuts and bolts of the statistical analyses than do other chapters in this volume. For an introduction to the basics of ANOVA, see Chapter 3. Although we focus on competition, many of the points raised are equally applicable to experiments on other types of species interactions such as predator–prey relationships or mutualisms.

4.2 Ecological Questions about Competition

The simplest question we can ask about competition is whether or not it occurs in the field. To answer this question, it is necessary to have experimental treatments in which the absolute abundance of potential competitors is manipulated and to test whether organisms in the treatments with lower abundance of potential competitors perform better. The difference in performance between such abundance treatments is the magnitude of competition (or magnitude of facilitation if

performance is better with higher abundance). Finding that competition does or does not occur is an important preliminary step in any field investigation of competition but, by itself, is not very informative. Most of the important questions about competition involve comparisons of the magnitude of competition and therefore involve more complex experimental designs and analyses than simply a comparison between two or more abundance treatments (Goldberg and Barton, 1992).

One group of questions requires the comparison of the magnitude of competition among environments (sites or times). For example, field observations might suggest the hypothesis that the distribution of a species is determined by the sum of competition from all other species at the same trophic level. A field experiment to test this hypothesis must compare the magnitude of competitive effects on that focal species in sites where it is abundant with sites where it is absent or rare (e.g., Hairston, 1980; Gurevitch, 1986; McGraw and Chapin, 1989). Similarly, to resolve the current controversy among plant ecologists over whether the importance of competition increases with increasing productivity or stays constant, it is necessary to compare the magnitude of competition on populations or communities among sites that differ in productivity (e.g., Wilson and Tilman, 1991; Campbell and Grime, 1992). Statistical analysis of an experiment to compare magnitudes of competition among sites is described in Section 4.4.1.

A second group of questions requires comparison of the magnitude of competition among taxa, i.e., comparisons of competitive ability. For example, classical competition theory predicts that, for coexisting species, intraspecific competition is greater than interspecific competition. Mechanistic models of competition make predictions about particular traits that are related to competitive ability (e.g., Grime, 1977; Schoener, 1986; Tilman, 1988; Werner and Anholt, 1993) and these require comparisons between taxa that differ in these traits. Similarly, quantifying the magnitude of selection on different traits due to the competitive environment requires a comparison of the magnitude of competition among different phenotypes or genotypes. Statistical analysis of experiments to test hypotheses about rankings of competitive abilities and relationship of competitive abilities to traits of organisms is described in Section 4.4.2.

Finally, some questions require comparisons of both environments and taxa. For example, many of the mechanistic models that generate predictions about traits related to competitive ability also make predictions about the ways in which those traits change between environments or about tradeoffs between competitive ability and response to other processes such as predation or disturbance (see references above). These predictions can be tested by comparing the magnitude of competition between different taxa in different environments (e.g., different resource availabilities, natural enemy densities, disturbance rates). These are necessarily highly complex designs with many independent factors, and the main points of biological interest will often be statistical interactions between factors (e.g., taxon × competition × environment interactions to test whether competi-

tive hierarchies change between environments). However, highly multifactorial designs will often also yield uninterpretable interaction terms as well as requiring enormous sample sizes; the addition of such complexity should be carefully considered in light of the ecological question of interest.

It is important to note that the basic competition experiment described above cannot by itself address the mechanisms underlying competitive interactions. Negative interactions can occur through direct interactions (interference competition), through a variety of shared, limiting resources (exploitation competition), through shared natural enemies (apparent competition), and other complex routes. Understanding these mechanisms is essential for developing generalizations about the role of competition in explaining evolutionary and ecological patterns (Schoener, 1986; Tilman, 1987). However, a broad diversity of approaches (field and laboratory, observational and experimental, manipulation of processes other than competition) is needed to examine specific mechanisms of interaction, making it difficult to provide a general discussion of experimental designs and statistical analyses. Therefore, the focus of this chapter will be on only one part of what is needed for a full understanding of competition: measurement and comparison of the magnitude of competition in the field.

To summarize, we emphasize that many of the important ecological questions about competition entail comparisons beyond simply demonstrating that competition occurs between a particular pair of species at a particular time and place. Therefore the first, and perhaps most important, recommendation of this chapter is that the first step in conducting a competition experiment be careful consideration of the goal of the experiment to identify appropriate comparisons. This may sound trivial and obvious but the literature is replete with statistically perfectly designed and analyzed experiments that do little to explain an observed pattern, to test assumptions or predictions of theory, or to provide a sound basis for management decisions in a particular system.

4.3 Experimental Design

4.3.1 Terminology

Before diving into the details of different types of experimental designs and the questions for which each is appropriate, we present some basic terminology. A *focal taxon* is one whose response to competition we are measuring. An *associate taxon* is one whose effect on the focal we are measuring, i.e., the one whose abundance is being manipulated. (Note that, in some experimental designs, the same species can be both a focal and an associate species.) A *background taxon* is one that is present in all experimental treatments but is not an explicitly designated focal or associate group. Background taxa can include other potential competitors, resources, natural enemies, or mutualists. "Taxon" in these three

definitions will often be species, but could also represent genotypes or groups of species (see Section 4.3.6).

A *response parameter* is the aspect of performance of the focal species that is measured. *Individual-level* responses include aspects of behavior, morphology, and physiology, and components of the fitness of individuals (e.g., growth rate, survival probability, or reproductive output). *Population-level* responses include population size or growth rate, where population size can be measured as density, biomass, cover, or other measure of abundance. *Community-level* responses include parameters such as taxonomic or functional group composition, degree of dominance, or diversity.

Competitive abilities can be compared either among focal species (*competitive response*) or among associate species (*competitive effect*) (Goldberg and Werner, 1983) . The distinction is important because different traits can determine ability to suppress other organisms (competitive effect) and ability to tolerate or avoid suppression (competitive response) (Goldberg and Landa, 1991). Competitive effects can be measured at the *natural abundances* of the associate or on a *per-unit amount* basis. Common measures of abundance are density, biomass, and cover (for sessile organisms), but other measures may also be appropriate (e.g., total root length or leaf surface area for plants). The comparison of results using different measures of associate abundance may be informative in itself. For example, the result that species have different per-capita effects but similar per-gram effects would indicate that the major trait influencing per-capita competitive effect is simply biomass per individual.

4.3.2 Basic Experimental Designs

Competition experiments fall into three general categories defined by the way in which density is controlled: substitutive designs, additive designs, and response surface designs (Fig. 4.1; Silvertown, 1987). For all categories, it is essential that densities are experimentally manipulated. If natural variation in abundance is used to test competition, environmental differences potentially correlated with the natural density gradient of the associate may also directly affect the focal individuals, thus confounding competitive effects with other environmental factors, whether biotic or abiotic.

In substitutive experiments (replacement series), total density is kept constant and frequency of each species is varied (Fig. 4.1A). Numerous critiques of substitutive experiments and constraints on its use have been published over the last decade or so and its use is not generally recommended for experiments in natural communities (see Connolly, 1986, and references therein). A substitutive experiment tests only for the relative intensity of intra- and interspecific competition. So, even if all of the rather restrictive assumptions of the design are met, it is never an appropriate design to test for the occurrence of competition or for comparisons of the absolute magnitude of competition. For questions that require only the relative magnitude of competition, however, such as whether or not

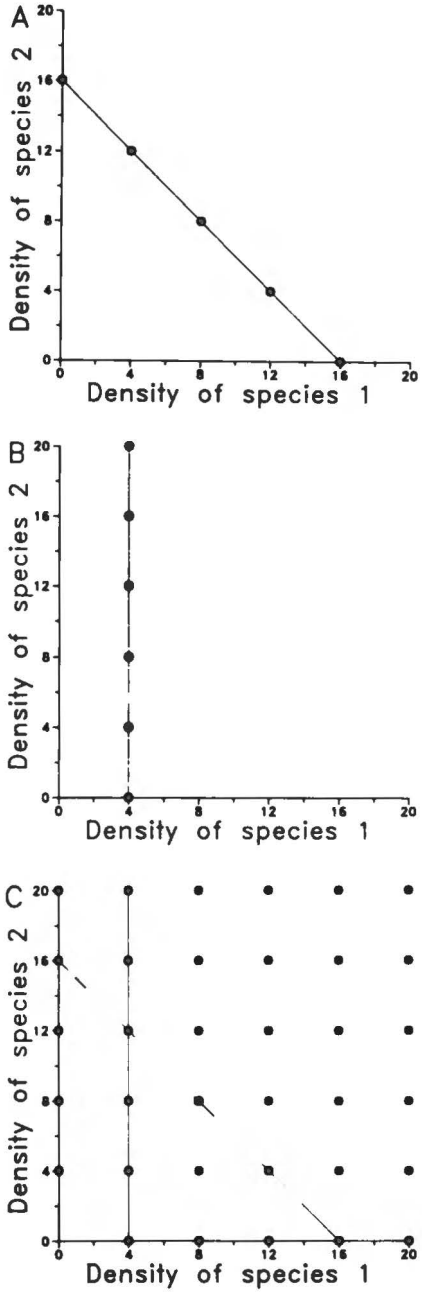

Figure 4.1. Three categories of design for competition experiments where densities of two species (or genotypes or groups of species) are manipulated: (A) substitutive experiments, (B) additive experiments, and (C) response surface experiments. Each point on the graphs represents a single experimental treatment. In (C), the subsections of the phase plane that constitute the entire experiment in (A) and (B) are shown as solid lines.

niche separation occurs, substitutive designs can be useful if the assumptions are met (see Connolly, 1986).

In additive experiments, density of the focal species is kept constant and density of the associate species is varied experimentally (Fig. 4.1B). Although usually described with only two species but many density treatments, this definition actually fits many field experiments on competition. For example, "removal experiments" typically compare response of a focal species between a treatment in which the associate is present at its natural abundance and a treatment where the associate is completely removed. If the focal species is at a constant density, this is a form of additive design with only two associate densities: presence and absence. Keeping the focal density constant in all treatments is important as this keeps the level of intraspecific competition constant. For some situations, this may mean either removing or introducing some individuals of the focal species into experimental plots. If focal individuals are added, it is critical that all focals in all treatments be manipulated the same way, i.e., all introduced or all naturally occurring individuals. Section 4.3.5 describes several important considerations in choosing associate density treatments. The major limitation on additive designs is that they confound density and frequency effects. As density of the associate increases, it also increases in frequency, i.e., it represents a greater proportion of the total mixture of the two species because the focal species is kept at a constant density. If the magnitude of competition is frequency as well as density dependent, this can be a serious problem.

The third design listed above, the response surface experiment (or addition series, sensu Silvertown 1987), gets around this problem by manipulating densities of both focal and associate species (Fig. 4.1C). This type of experiment can provide the data needed to develop realistic population dynamic models (Law and Watkinson, 1987). Although this is the ideal two-species design, the large number of density combinations needed makes such an experiment impractical for many field situations, especially when several species combinations must be investigated. Even in the laboratory, relatively few experiments have explored large combinations of densities and we know of no field experiments that have used it. Therefore, we will not address analysis of these experiments further; the interested reader should see Ayala et al. (1973) and Law and Watkinson (1987).

4.3.3 Level of Response Variables and Problems of Temporal and Spatial Scale

Most field experiments on competition use individual-level response parameters such as foraging behavior or individual growth rate because of constraints on both the spatial and temporal scales of experimental manipulations. Only for small, short-lived organisms is it feasible to measure growth rates of entire populations. This constraint means that conclusions about the population consequences of competition for distribution or abundance can be inferred only from

individual-level measures rather than directly demonstrated. For long-lived organisms, this raises the question of which life history stage(s) of focal individuals is most appropriate to use. Ideally, experiments should be conducted with all ages or stages, in which case models of age- or size-structured population dynamics could be used to estimate population growth rates under various experimental conditions (e.g., Gurevitch, 1986). In practice, this is not always feasible and the subset of stages to be studied should be chosen based on their likely importance in population regulation. Sensitivity analysis of matrix population models of demographic data can be useful in determining the best ages or stages to study (Caswell, 1989; Chapter 11).

Relatively few field experiments have measured community-level responses to ask how species composition or diversity are affected by competition (Goldberg and Barton, 1992). Possible approaches for analysis of community-level response are to use relative population abundance as the response variable in univariate analyses like those described in Section 4.4, to use diversity indices that summarize community properties in such univariate analyses, or to use absolute abundances of all species in multivariate analyses of variance (Chapter 5).

4.3.4 Absolute vs Relative Response Variables: When to Standardize?

Almost any response variable can be expressed as an absolute value or as a percentage of the value of that variable in the absence of associates. Which of these is used in an analysis can have a major impact on the interpretation of comparisons of the magnitude of competition. Figure 4.2 shows an example in which the growth response of two focal species to a density gradient of the same associate species are compared. Using absolute values of focal response, species 2 has lower growth than species 1 in the absence of any associates (lower intercept) and also has a less steep slope. Thus, species 2 would be considered a better response competitor (less absolute decrease in growth for each associate individual added; Fig. 4.2A). However, when expressed as a percentage of growth in the absence of the associate (intercept for both is 100%), species 1 now shows less of a reduction in growth and so would be considered the better response competitor (Fig. 4.2B).

This potential conflict in interpretation between standardized and unstandardized data is most obvious for comparisons of competitive response of different focal species or of the same focal species in different environments (e.g., high and low productivity) because intercepts will usually be different with unstandardized data. However, it may also appear in comparisons of competitive effect among associate species. Normally, we would expect that a single focal species in a single site would have the same intercept (performance in the absence of any associates) in regressions with different associate species, so that standardized and unstandardized data would yield the same result. However, when the zero-density treatment is a result of experimental removal of the associates, two factors

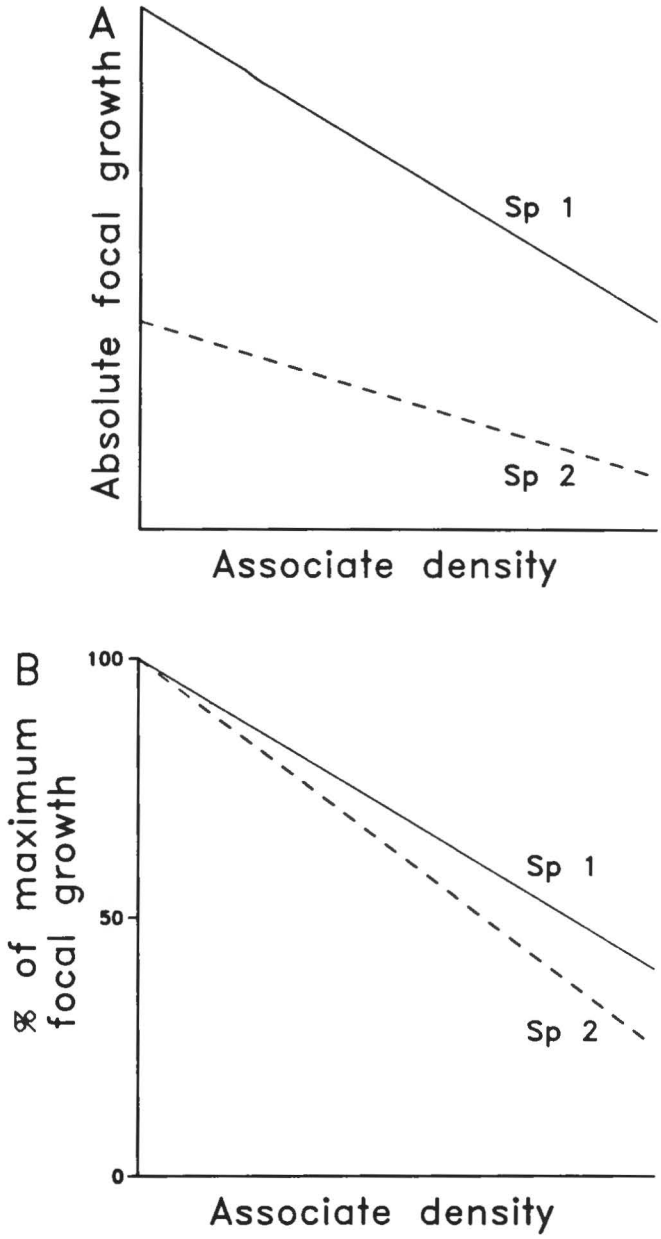

Figure 4.2. A hypothetical example of results of an additive experiment with two different types of response variables for two focal species: (A) absolute growth and (B) percentage of growth in the absence of the associate. Using absolute response, species 2 is the better response competitor, while species 1 is the better response competitor using percentage of maximum growth.

may lead to differences in intercepts among associate species. First, there may be residual effects of the former presence of the associate that could differ between associate taxa. This result is more likely to occur with sessile organisms that could have cumulative effects on a single location. For example, plants may have different residual effects on nutrient availability due to differences in litter quality and subsequent decomposition rates or soil organic matter accumulation. In addition, roots can seldom be removed and the presence of dead and decaying roots could also have significant residual effects. Second, associates can be removed only where they were initially present. If different associate species were initially present or abundant in different microenvironments and these microenvironmental conditions influence the focal species directly, focals may differ in zero-density treatments for different associates, i.e., differ in intercepts. Both of these potential problems point out the importance of having separate zero-density treatments for each of the associate species used in a removal experiment.

There has been remarkably little discussion in the literature about the conditions in which response in the presence of associates should be standardized to response in the absence of associates; the following recommendations should be viewed as preliminary. For questions concerning the consequences of competition for distribution and abundance, relative values are probably most useful. For example, suppose that competition reduces individual growth of a focal species by 10 g in an unproductive site and by 100 g in a productive site, suggesting that competition is more intense in the more productive site. However, if the growth in the absence of competition were 10 and 200 g in the unproductive and productive sites, respectively, the relative responses are 0% and 50%, suggesting competition is more intense in the unproductive site and, in fact, intense enough there to exclude the focal species. [See Campbell and Grime (1992) for an example of contrasting results for standardized and unstandardized data in a competition experiment.]

On the other hand, if the mechanisms of competition are under investigation, absolute values may be more useful. For example, it may be of interest to relate reductions in resource availability due to consumption by an associate species to changes in resource consumption and growth in the focals. Using standardized data would make it difficult to understand the processes underlying these relationships. Absolute responses are also necessary when competitive effects and responses are being quantified to use as input into models of population interactions because most population dynamic models use absolute, not standardized, parameters. For example, in the Lotka-Volterra competition equations, equilibrium density in the absence of competition (K) is an explicit parameter and both K and the per-capita competition coefficients (α_{ij}) have an impact on the dynamics and equilibrium outcome of the interaction.

4.3.5 Manipulation of Competition Intensity: How Many Density Treatments?

The range of density variation in an additive experiment could be simply presence vs absence of the associate or several densities could be experimentally imposed.

Presence/absence comparisons are appropriate where the total magnitude of competitive effect under natural conditions is all that is required. However, unless competitive effects are completely linear, presence/absence comparisons (or comparisons of any two densities) do not allow accurate calculation of per-capita effects (Fig. 4.3). Where the density dependence of per-capita competitive effects has been tested, it is rarely constant (Harper, 1977; Schoener, 1986; Goldberg, 1987; Miller and Werner, 1987), and so interpolating from presence/absence comparisons to per-capita effects across a range of densities is probably not generally safe. Comparisons of competitive effect between associate species and,

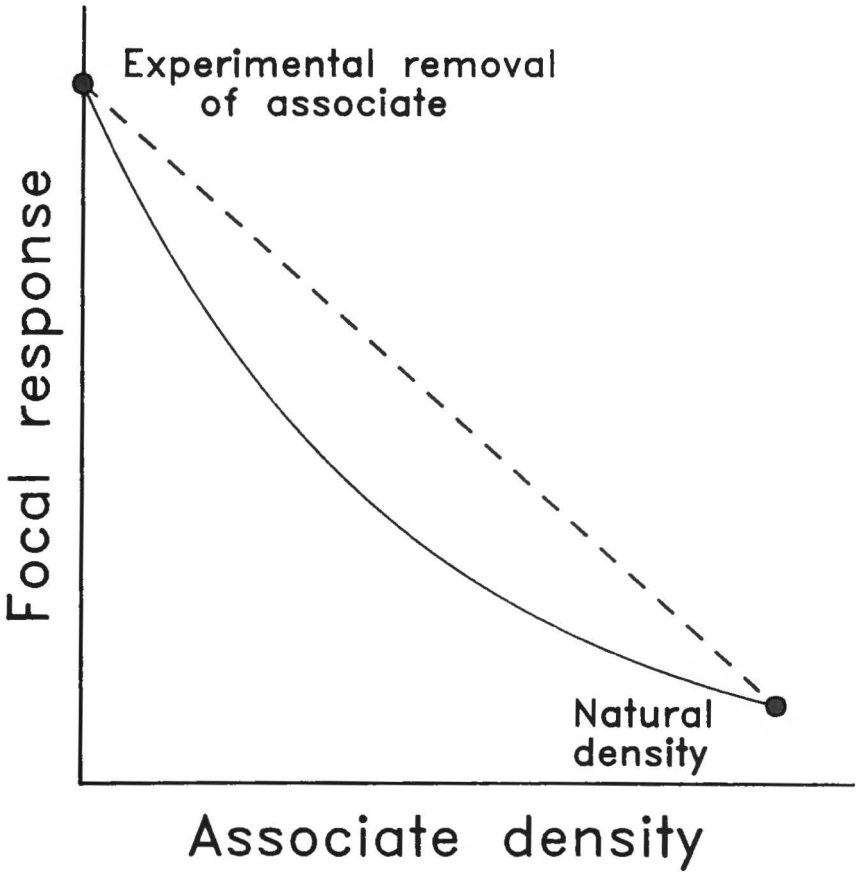

Figure 4.3. An example of inaccurate estimation of per-capita competitive effects from an experiment with two treatments: the associate present at its natural density and the associate completely removed. The true per-capita effect is shown as a solid line. A linear extrapolation from the two treatments (dashed line) underestimates the true per-capita effect at low associate density but overestimates the true effect at high associate density.

especially, relating these competitive effects to traits requires per-capita effects so that effects of total density and per-capita effects (or, more generally, total abundance and per-unit amount) are not confounded.

If multiple densities are chosen, many density treatments with low or no replication will usually be better than few, well-replicated densities. The appropriate analysis will then be a comparison of slopes from regressions of the response variable on density rather than contrasts between means of different treatments (see Section 4.4.2 and Chapter 8). As both the strongest nonlinearities and the greatest variation among replicates tend to be at the lowest densities, concentration of density treatments at the low end rather than an even distribution is recommended. Densities above natural are also useful because (1) naturally occurring densities usually fluctuate considerably and (2) competition might occur only at densities above those normally determined by other factors such as predators. Regardless of the range of densities used in an experiment, it is always important in publications to identify the natural density range so readers can assess the applicability of results to natural conditions.

4.3.6 Pairwise vs Total Competition: Which Associate Groups to Use?

The associate group whose abundance is being manipulated can be a single species or a group of species considered together. Which is more appropriate depends on both the biology of the species under study and on the goal of the experiment. The tendency has been for animal ecologists to measure competition between pairs of species chosen a priori as likely strong competitors while plant ecologists have more often simply quantified total competition from the entire plant community on a focal species. Because plants all require the same few resources, all plants are potential competitors and there is often little a priori basis for choosing subsets. Therefore the most common type of field competition experiment in plant communities is a presence/absence additive design where the associate "species" is the entire plant community except for the focal individuals (Goldberg and Barton, 1992). For many animal species, on the other hand, more restricted guilds of potentially interacting species that share resources can often be defined (e.g., seed-eaters or insect-eaters). The danger here is that while potential competitors are often chosen by taxonomic relatedness (e.g., congeners), distantly related taxa may be just as likely to share resources and compete for them (e.g., seed-eating rodents, ants and birds in deserts; Brown et al., 1986).

When the question of interest concerns comparisons of competitive ability among species and relationships of competitive ability to traits, pairwise experiments with single species as associates and focals are usually most appropriate. The number of species also becomes important because tests of these relationships essentially use each associate or focal species as a "replicate" in a secondary statistical analysis. Use of only a single species with each value of a trait is unreplicated for tests of hypotheses about species characteristics related to com-

petitive ability. For example, if only two associate species are compared and the one with larger leaves has a stronger negative effect, the conclusion that leaf size confers strong competitive effect is unwarranted because of lack of replication within each leaf size class. The taxa may also differ in other correlated traits such as root length or nutrient uptake rate. As leaf size is confounded with these other variables, assigning a causal effect to leaf size alone is not justified. We recommend either replicating species within trait classes or using species with a range of values of a continuous trait.

4.3.7 Direct Effects, Indirect Effects, and Higher-Order Interactions: What to Do with Background Species?

Response of a focal species to manipulation of abundance of an associate species represents the sum of the direct effects of the associate plus all the indirect effects and higher-order interactions involving the associate and all the background species, the *net effect*. If no background species are present, only direct effects are measured (but see below). *Indirect effects* are defined as changes in the effect of an associate on a focal species when background species change the abundance of the associate (linkage indirect effects, sensu Miller and Kerfoot, 1987). *Higher-order interactions* are defined as changes in the effect of an associate on a focal species when a background species changes the per-capita effect of the associate (behavior indirect effects, sensu Miller and Kerfoot, 1987).

If the primary question concerns the effect of the associate species on the distribution and abundance of the focal species, net effect may in fact be what is desired. For example, suppose removal of a dominant has little effect on a focal species because other, unmeasured, species respond to its removal more quickly than does the focal species and thus end up suppressing the focal species just as much as did the designated associate. The conclusion that competition from that associate species is unimportant in controlling the abundance of that focal species is reasonable, even though the associate may have strong direct effects on the focal in a pairwise experiment with no background species. However, if the primary question is which traits of an associate species allow it to become dominant, measuring its net competitive effect will be misleading because traits of the background species will also influence the results.

Therefore our general recommendation is that questions about the relationship of traits to competitive ability or determination of competitive hierarchies are usually best addressed in pairwise experiments that minimize the potential for indirect effects or higher-order interactions. In contrast, questions about the consequences of competition for abundance and distribution are usually best addressed by incorporating any nondirect effects into the response variable, i.e., by leaving background species in all treatments. We point out, however, that strong nondirect effects restrict generalization across sites or times even more than is usual in ecological field experiments because the biotic environment and

therefore complex of nondirect effects differ as well as the abiotic environment and history of the sites.

In reality, almost any field experiment of any design has background species and therefore we can never measure only direct effects. It is rarely possible or even desirable to eliminate completely all the biota except for the designated focal and associate species. First, exploitation competition itself is an indirect interaction, mediated through shared food resources. When the resource is a living organism, even if the food is not an explicitly designated associate or focal species, it obviously must be present in the experiment. Similarly, negative interactions may be mediated by shared natural enemies (apparent competition; Holt, 1977; Connell, 1990) or mutualists, and removing "background" species would preclude detecting an important interaction between two "competitors" (but appropriate if quantification of only exploitation competition is desired). Monitoring the dynamics of these intermediary nonmanipulated species, whether resources, natural enemies, or mutualists, can be an important step toward understanding the mechanisms of interactions. Second, microorganisms are part of the biota and may be important links in chains of indirect effects, but are very difficult to remove completely. The important point in terms of designing and interpreting field experiments is that it is critical to recognize which background species are present and to incorporate this knowledge when interpreting results.

4.4 Statistical Analysis and Interpretation of Results

4.4.1 Comparison of Habitats: Effects of Competition on Distribution and Abundance

One of the most common questions about competition is whether it influences the distribution and abundance of a focal species. The recommended approach to this question is to repeat a presence/absence additive design with background species in differing habitats, those where the focal species is present or abundant and those where it is absent or rare. This question is really a population-level one and ideally population-level response parameters should be used (Section 4.3.3). A minimum of two replicate sites in each habitat category (focal abundant vs rare) should be used so that the relevance of habitat differences to focal species distribution can be inferred from the results. With only a single site in each habitat category (admittedly the most common design actually used), the only valid inference that can be made is whether sites differ in the magnitude of competition. That is, site is confounded with habitat.

The choice of associate species can be a single species chosen a priori, a subset of species, or the entire community (see Section 4.3.6). Since the question concerns the net consequences of the presence of some associate species or group of species, it is appropriate to leave all other species present as background

species in all treatments, i.e., to quantify the sum of all direct, indirect, and higher-order effects.

The actual experiment would involve establishing an experimental area in each site within which the focal species would be added and the associate species removed as appropriate. It is important to add individuals of the focal species back into its native habitat as well as the habitat from which it is absent to account for transplanting effects. Such a design also ensures that the density of the focal species is the same across all treatments (Section 4.3.2).

For an experiment with a single focal species, a single associate species or group, and several sites of each habitat type, the appropriate ANOVA model is

$$X_{ijkl} = \mu + \tau_i + \theta_j + \tau\theta_{ij} + \Sigma(\theta)_{jk} + \tau\Sigma(\theta)_{ijk} + \varepsilon_{ijkl} \qquad (4.1)$$

where μ is the overall mean, τ_i is the deviation due to the ith associate treatment (present at natural abundance or removed), θ_j is the deviation due to the jth habitat (focal species naturally abundant or rare), $\tau\theta_{ij}$ is the deviation due to the interaction of the ith treatment with the jth habitat, $\Sigma(\theta)_{jk}$ is the deviation due to the kth site nested within the jth habitat, $\tau\Sigma(\theta)_{ijk}$ is the deviation due to the interaction of the ith treatment with the kth site nested in the jth habitat, and ε_{ijkl} is the deviation due to the lth replicate within each site-treatment combination. Treatment and habitat are both fixed factors and site is a random factor, so the overall model is a mixed cross-nested design. The degrees of freedom, estimated mean squares E(MS), and F tests for this model are shown in Table 4.1 and SAS commands in Appendix 4.1A. See Searle (1971) for a detailed explanation of the underlying statistical theory. More complex designs can also be used that include blocking factors within sites (Chapter 3).

The important result from this ANOVA is whether the treatment × habitat interaction is significant, i.e., the magnitude of competition differs between habitats where the focal species is abundant and where it is absent or rare. The expected pattern of results is shown in Fig. 4.4, with a larger magnitude of competition where the focal species is absent or rare than where it is abundant. Notice that, in this analysis, the treatment × habitat interaction is tested over the treatment × site interaction and the power of the test is a function of the number of sites. Thus, the most powerful design to use in this case is to maximize the

Table 4.1 ANOVA for a Completely Balanced Cross-Nested Design

Effect	df	E(MS)	F-ratio
Treatment	$t-1$	$\sigma^2_e + n\sigma^2_{\tau\Sigma(\theta)} + hsn\sigma^2_\tau$	$MS_T/MS_{TS(H)}$
Habitat	$h-1$	$\sigma^2_e + tn\sigma^2_{\Sigma(\theta)} + stn\sigma^2_\theta$	$MS_H/MS_{S(H)}$
Treatment × habitat	$(t-1)(h-1)$	$\sigma^2_e + n\sigma^2_{\tau\Sigma(\theta)} + sn\sigma^2_{\tau\theta}$	$MS_{TH}/MS_{TS(H)}$
Site (habitat)	$h(s-1)$	$\sigma^2_e + tn\sigma^2_{\Sigma(\theta)}$	$MS_{S(H)}/MS_e$
Treatment × site (habitat)	$(t-1)h(s-1)$	$\sigma^2_e + n\sigma^2_{\tau\Sigma(\theta)}$	$MS_{TS(H)}/MS_e$
Residual	$ths(n-1)$	σ^2_e	

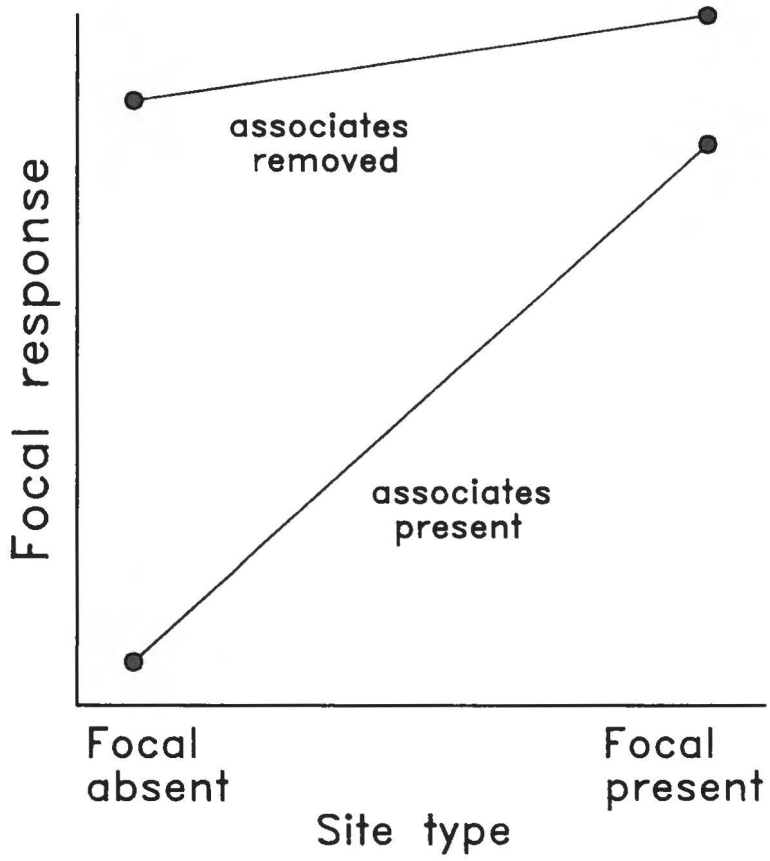

Figure 4.4. Pattern of predicted results from a presence/absence additive design if competition influences the distribution and abundance of the focal species. The key result is that the magnitude of competition (depression due to presence of competitors) is much greater where the focal species is normally absent than where it is normally present (i.e., a habitat × competition interaction). Note that, in this case, the focal species also performs better in the habitat where it is abundant than where it is absent even in the absence of associates and competition is present in both types of habitats so that both habitat and competition would probably be significant main effects.

number of sites within habitat types and minimize the number of replicates within sites. Obviously, such a strategy creates the potential for logistical difficulties if the sites are any distance apart.

When the number of replicates is not the same for each level of all treatment factors, the experimental design is termed "unbalanced" (Searle, 1987). Unbalanced designs are very common in ecological experiments because of insufficient material (e.g., it may be more difficult to get replicates from one site or species

or genotype than others) or loss of replicates through mortality. Unbalanced ANOVA designs are more difficult to analyze and, unlike balanced designs, there is no single correct way to analyze them. Readers may wish to consult recent discussions of the analysis of unbalanced designs (Milliken and Johnson, 1984; Fry, 1992; Shaw and Mitchell-Olds, 1993). SAS 6.0 and above will construct an approximate *F*-test for unbalanced designs using the Satterthwaite approximation for the degrees of freedom (Satterthwaite, 1946; Hocking, 1985; see Appendix 4.1A). If SAS procedure GLM is used for the analysis, Type III sums of squares (SS) must be requested; the *F*-tests provided by the other types of SSs (I, II, and IV) are incorrect. Because there are several approaches to this problem and no general agreement, the individual researcher must evaluate whether the approach taken by SAS is appropriate for the particular design employed (see discussion in Chapter 3).

A new alternative approach to the analysis of unbalanced designs that contain both fixed and random effects—mixed models—is maximum likelihood (McLean et al., 1991), which is implemented in the latest version of SAS (6.07, SAS Institute Inc., 1992) as procedure MIXED (Appendix 4.1B). One potential advantage of this method is that it will likely prove to be more robust to unbalanced designs than the standard least-squares approach, although this has not yet been proved. A disadvantage is that computations for complex designs can take a long time and sometimes will fail to converge on a solution. Be warned that the documentation of the technique is not completely clear, especially for those without extensive statistical training; consultation with a statistician is highly recommended.

4.4.2 Comparison of Species: the Relationship of Traits to Competitive Ability

Comparing competitive abilities among species (or other groups) and determining the relationship of competitive abilities to traits will usually require per-capita measurements. The recommended approach to compare species is an additive design with multiple densities of the associate, using either a single focal species and several associate species that differ in value of some trait (comparison of competitive effect) or a single associate species and several focal species (comparison of competitive response; Fig. 4.2A). A significant relationship (nonzero slope) between density of the associate and response of the focal species indicates competition (negative slope) or facilitation (positive slope).

The analysis is stepwise (see Zar, 1984, Fig. 18.1). The first step is to test for differences in density effects among focal or associate species, a test for homogeneity of slopes. The ANCOVA model is

$$X_{ijk} = \mu + \Delta_i + \alpha_j + \Delta\alpha_{ij} + \varepsilon_{ijk} \tag{4.2}$$

where μ is the mean intercept (the value of the response parameter at 0 density), Δ_i is the deviation due to the ith density treatment (the average slope of the line

that predicts the effect of associate species density on the response parameter), α_j is the deviation from the mean intercept due to the jth associate or focal species (differences in intercepts among associate or focal species), and $\Delta\alpha_{ij}$ is the deviation due to the interaction of species identity and density (differences in slopes among species; Fig. 4.2A). Δ_i is a continuous variable and α_j is a categorical factor.

In this first analysis, we are concerned only with the interaction term. We ignore the main effects of density and species identity at this juncture because such main effects are meaningful only if the interaction effect is declared not statistically significantly different from 0. The SAS commands for performing this analysis are given in Appendix 4.2. If differences among slopes are found, then a post-hoc Tukey-type multiple comparison test can be performed (Zar, 1984, Section 18.6) to discover which particular species differ from each other. Alternatively, if one has a priori hypotheses concerning differences between particular species or sets of species, then a series of specified comparisons (contrast tests) should be done rather than a single overall ANCOVA.

A potential problem in this analysis arises if only a single zero-density treatment (no competitors) is used for all associates, because it is impossible to assign a treatment level (identity of associate) to replicates at this density. In this case, it is reasonable simply to randomly allocate all of the zero-density replicates among the associate species. (For example, if there are three associate species and 30 zero-density replicates, then 10 replicates would be randomly assigned to each species treatment.) This problem should arise only when all associate individuals are added (e.g., to cages or in greenhouse experiments), because in field removal experiments, there will usually be distinct zero-density treatments for each associate species (see section 4.3.4).

If slopes are homogeneous, one can then proceed to (1) measure the average slope (Δ_i) and (2) test for differences in intercepts (α_j) among species using the following model:

$$X_{ijk} = \mu + \Delta_i + \alpha_j + \varepsilon_{ijk} \tag{4.3}$$

Note that the model is identical to that in Eq. (4.2) except that the interaction term has been deleted. The SAS commands are given in Appendix 4.3. Significant density effects (a nonzero slope) indicate that competition or facilitation is occurring and the average slope measures the general per-capita effect of all associates on the focal species.

In comparisons of focal species, significant species effects (α_j) indicate simply that the focal species differ consistently across all densities of the associate species (e.g., in maximum growth rates or sizes). In comparisons of associate species, significant species effects are more problematic because the focal species remains the same and thus should not differ in response to the associates when no associates are present. Some biological interpretations of significant associate

main effects were given in Section 4.3.4. An alternative interpretation of either significant focal or associate effects is that it could be an artifact of using linear regressions on nonlinear data. Figure 4.5 shows an example where two focal species have the same slopes but different intercepts in linear regression analyses. However, inspection of the graphs shows that the intercepts are actually very similar but that species B is strongly nonlinear at low associate densities. If

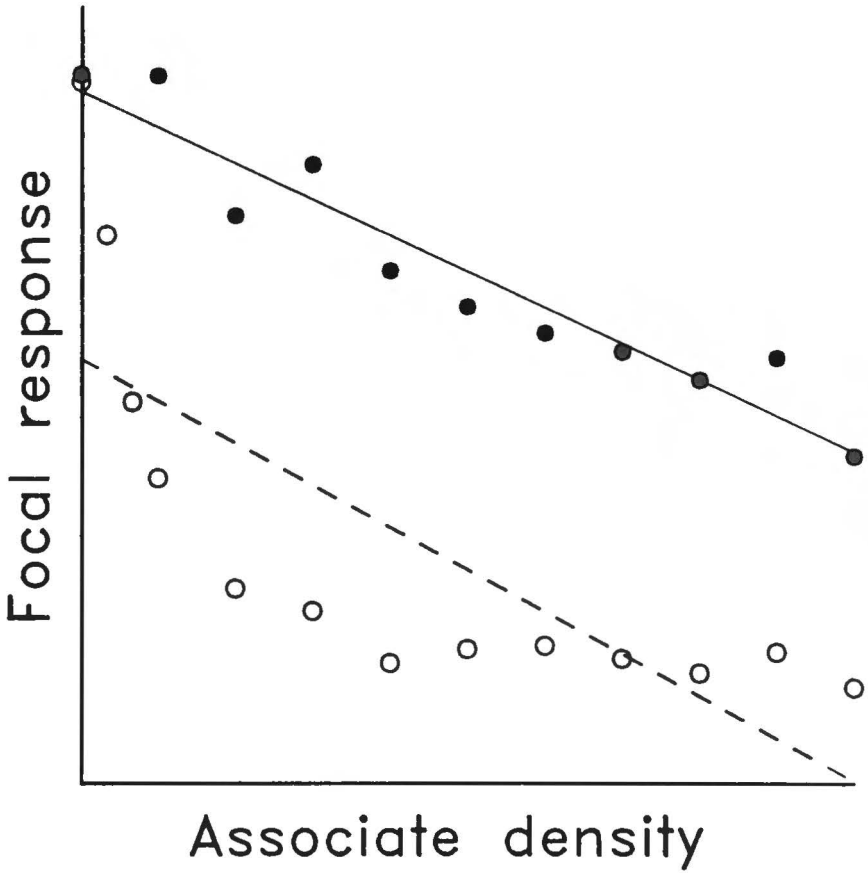

Figure 4.5. An example of misleading results of comparison of slopes (individual competition coefficients) using linear regression models. Using linear regressions, the two species would appear to have different intercepts, but similar slopes, i.e., similar competitive effects if the curves represent two associate species. However, inspection of the data reveals the intercepts are actually very similar and species B (open circles) has a strongly nonlinear effect on the focal species at low densities. Inspection of the residuals of the linear regression for species B would show a nonrandom distribution with associate density.

transformations do not succeed in linearizing the data, a nonlinear regression analysis is warranted (Chapter 8).

Relating per-capita effects or responses to values of a particular trait can be conducted as a second-level statistical analysis by correlating slopes with trait values (e.g., Goldberg and Landa, 1991). With fewer species or with few, discrete values of the trait, slopes can be compared with a priori contrasts among groups of the associate or focal species (Day and Quinn, 1989) or in the context of the ANCOVA where species are nested within species types in a way exactly analogous to sites nested within habitats (see Eq. 4.1).

4.4.3 The "Envelope Effect"

The performance of an organism is likely to be affected by many factors (e.g., genotype, weather, predators) besides the density of competitors. Usually in field experiments such additional factors are either poorly controlled or not controlled at all and effects of competitors may be obscured by the large amount of variation caused by these other factors. This general problem may arise in any type of analysis of field experiments; in regression analyses of competition, it may be manifested as an "envelope effect" (Firbank and Watkinson, 1987; Goldberg, 1987; Fowler, 1990a). Competition may act to keep the response of the focal species below some maximum, while within this competitively determined "envelope" other factors act to further depress the focal species. The result is that a plot of the response variable against density will be a scatter of points forming a triangle in the lower left-hand portion of the graph (Fig. 4.6).

If envelope effects occur, they can cause considerable difficulty in quantifying density effects. First, the effect will be obscured and the slope of the line decreased by the large amount of variation caused by other factors. Second, the estimated per-capita effect will be substantially less than the true effect because a simple linear regression will place the regression line through the middle of the points, rather than along the upper edge of the envelope. Because the variance is smaller at higher competition intensities, the mean will be closer to the upper edge of the envelope at higher competition intensities than at low intensities, thus lowering both the intercept and the slope. Third, the important statistical assumption of homoscedasticity, that the variation around the regression line is the same everywhere, is being violated. Clearly, the variation decreases with increasing density. In this case, there may very well be no transformation of the data that will make the variance homogeneous. Pacala and Silander (1990) developed one solution to this problem. They note that the residuals (the variation around the regression line) of such a distribution are distributed as a gamma function, which has the properties of being unimodal, left-skewed, and bounded at zero (see Chapter 12 for a description of the gamma function). In their experiment, competitive effects followed a hyperbolic function. Thus, rather than the more familiar

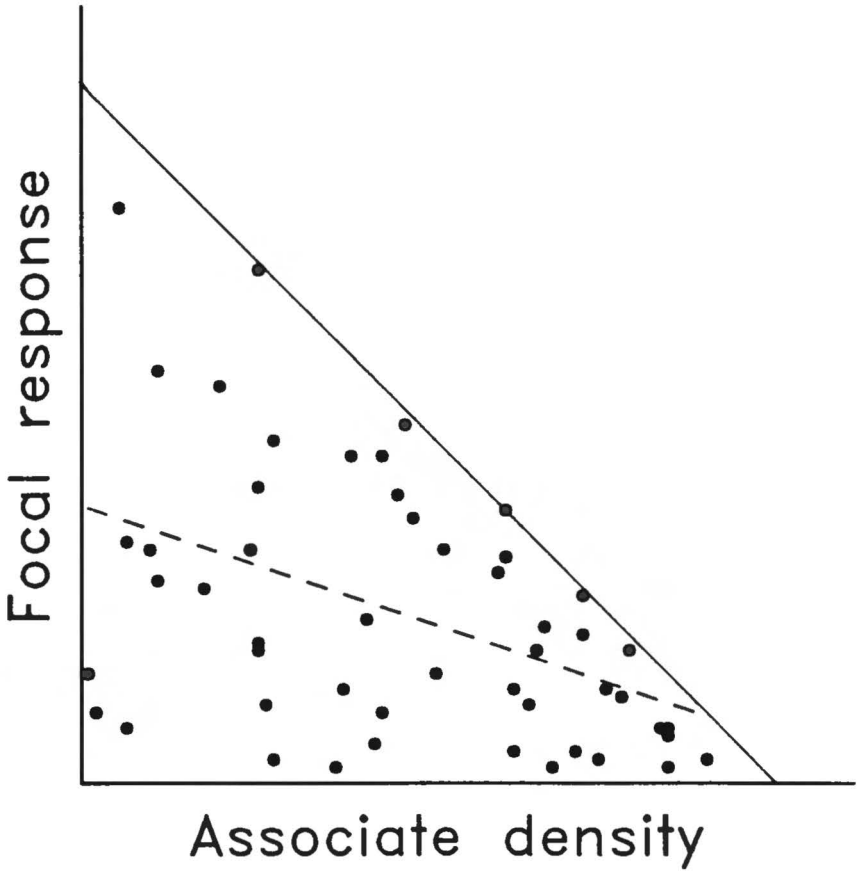

Figure 4.6. Example of the "envelope" effect where the variance in focal species response is much greater at low associate densities than at high associate densities. The solid line at the upper edge of the envelope represents the constraint on maximum focal response imposed by competition. The dashed line represents the competitive effect estimated by a standard least-squares linear regression.

least-squares method, they used a maximum likelihood estimator and iterative Newton method as described in Chapter 8.

Another possible solution to the problem of the "envelope effect" is to monitor other potentially interacting factors and use them as covariates in a multiple regression (Chapter 9). The possible drawbacks of this approach are (1) the logistical difficulty of measuring many additional variables, (2) the potential failure to measure the key factor, and (3) the fact that multiple regression brings with it an additional host of statistical assumptions and difficulties (Chapter 9). A final possible solution is to manipulate or hold constant other variables. Here

the drawback is the movement away from as natural a setting as possible, the entire reason for doing a field experiment to begin with.

4.5 Related Techniques

4.5.1 Isotonic Regression

An alternative and intermediate approach to the ANOVA and ANCOVA models presented above is isotonic regression, which can be thought of as a form of ANOVA to be used when there are ordered expectations (Barlow et al., 1972). This technique has been little used up to now in biology. Gaines and Rice (1990) present a cogent description of this technique and offer a microcomputer program for implementing it. The advantage of isotonic regression for detecting density-dependent responses is twofold. First, there are no assumptions as to the shape of the density response except that it is monotonic. For this reason, isotonic regression can be more powerful than standard least-squares regression analysis when the shape of the response is irregular or a step function. Second, fewer density treatments are needed than in ANCOVA because the precise shape of the density effect is not of interest. Instead, replication efforts can be placed elsewhere in more associate or focal species, more sites, or more replicates per site.

However, there are also several limitations to isotonic regression that restrict its usefulness to relatively simple questions about competition. First, per-capita effects cannot be calculated so it is not appropriate if comparisons are needed of species effects independently of their total abundance. Second, the technique is currently developed for analysis of different levels of a single factor. Thus, it could be useful in detecting the occurrence of negative effects of increasing density of a single associate but not in comparing associates or detecting interactions between competition and other factors. However, the method can be used in experiments with one additional factor other than level of competition if the dependent variable is *relative* magnitude of competition. In this special case, competitive effects are included in the response variable rather than as distinct treatments (section 4.3.4). For example, the reduction in response in the presence of competitors at their natural abundance relative to their complete absence could be compared over a series of sites with different expectations of the magnitude of this relative reduction.

4.5.2 Spatial Heterogeneity

This chapter has not dealt with the analysis of spatial heterogeneity except when spatial variation is an explicit part of the ecological question of interest (e.g., variation in magnitude of competition between sites or habitats). Chapter 3 discusses ANOVAs with block designs that explicitly account for variation due to spatial heterogeneity. Chapter 14 discusses an alternative approach to analysis

of spatial heterogeneity using variograms. Both of these methods require replication within a treatment, which is impossible using the multiple, unreplicated density design recommended in this chapter when comparisons of per-capita competitive abilities are required. There is currently no satisfactory general method that can both account explicitly for spatial heterogeneity and maximize investigation of potential nonlinearities. Therefore, the individual investigator must decide which is more important in the study and sacrifice number of blocks (replicates per density) or number of distinct densities accordingly. If an unreplicated regression design is chosen and the important environmental variables are known, these can be measured in individual experimental units and used directly as covariates in an ANCOVA. A third possible approach for incorporating spatial heterogeneity effects in designs with unreplicated densities is the permutation method described in Chapter 15, although this method is as yet untried for this application.

4.5.3 Beyond Simple Dependent Variables

The analysis of competitive effects potentially involves examining the response of many different types of response variables. The previous discussion has been in the context of examining a single variable that most likely meets the standard normality assumptions of parametric statistics. Time-based response variables such as survival time or time to emigration, while still single measures, often do not meet standard normality assumptions and require different analytic methods (Chapter 12).

Often the response variable is more complex than a single measure of focal performance. If individual-level measures are being used, for plants one might measure vegetative growth, seed number and seed size. If community-level measures are being used, one might measure the densities of several focal species. If the response variables (traits or species) are likely to be correlated with each other or to interact with each other, then a multivariate analysis of variance is appropriate (Chapter 5). If data are collected over time, for example, as growth curves of individuals or populations, then a repeated-measures analysis should be used (Chapter 6). For long-term experiments, time series analysis may be most appropriate (Chapter 7).

For all of these techniques, the basic design criteria outlined above are still appropriate, although different statistical techniques will be most powerful for different numbers of treatment levels vs replicates within treatments. In all cases, careful thought should be given to the type of analysis to be used at the time of experimental design, rather than waiting until the experiment is complete.

4.6 Some Final Comments on Experimental Design

We have argued that answering questions besides "does competition occur?" is critical to developing and testing general models of competition and that answer-

ing these questions will require more complex experimental designs than have often been used in field competition experiments. Nevertheless, there is a real danger that our experimental designs will become so complex that insufficient resources are available to detect any but enormous effects. Decisions on allocation of resources will always depend on the details of any particular study, but a few preliminary steps will aid in the process. We strongly recommend conducting pilot experiments in the field to estimate variances and thus the replication needed to detect effects of a given magnitude (Zar, 1984); results of such power tests may quickly demonstrate that experiments for some questions are simply logistically impossible in that system. If several factors are experimentally manipulated, replication should be focused on the factor(s) that will be used as the error variance(s) (denominator MS) for the most important test(s). As the example in Table 4.1 showed, this is not always at the level of replication of the smallest experimental units or the most detailed treatment combination (e.g., field plots). It is critical to work out the estimated mean squares for any design before beginning an experiment so that replication will be done at the proper level. Brownlee (1965) and Searle (1971) present examples for many common designs; consultation with a statistician at this stage is also highly recommended. The important point is simply that there will always be tradeoffs between the complexity of questions that can be asked from an experiment and the statistical power available for testing particular effects. Which of these is emphasized and which sacrificed is the most important decision underlying any experimental design.

Acknowledgments

We thank Brad Anholt, Jessica Gurevitch, and Earl Werner for valuable discussions of many of the issues raised in this chapter and Yoram Ayal, Drew Barton, Norma Fowler, Jessica Gurevitch, Ariel Novoplansky, and Earl Werner for their helpful comments on earlier drafts. The National Science Foundation provided support for the research on which this chapter was based. This collaboration was made possible by the telecommunications revolution.

Appendix 4.1

SAS program code for analysis of the cross-nested ANOVA shown in Eq. (4.1):

A. Analysis Using Least Squares

```
PROC GLM;
  CLASS T H S;                          /T = associate treatment, H = habitat,
                                          S = sites within habitat/
  MODEL X = T H T*H S(H)          /use type III sums of squares. Warning: results are
        T*S(H)/SS3;                     only approximate if the design is unbalanced./

*** If the design is unbalanced use the following command to
    obtain the correct F tests for the main effects ***

  RANDOM S(H) T*S(H)/TEST;    /Declare sites to be a random effect and ask SAS to
                                  compute the Satterthwaite correction for an
                                                        unbalanced design./

*** If the design is balanced use the following commands to
    obtain the correct F tests for the main effects ***

  TEST H=T E=T*S(H);                    /test associate treatment over
                                          the treatment-site interaction/
  TEST H=H E=S(H);                                /test habitat over site/
  TEST H=T*H E=T*S(H);                  /test treatment-habitat interaction
                                        over the treatment-site interaction/
```

B. Alternative Analysis Using Maximum Likelihood

```
PROC MIXED;                             /mixed model analysis using reduced
                                          maximum likelihood (the default)/
  CLASS T H S;
  MODEL X = T H T*H;                    /only fixed effects are placed
                                                  in the model statement/
  RANDOM S(H) T*S(H);           /random effects are specified separately/
```

Appendix 4.2

SAS program code for analysis of the ANCOVA model in Eq. (4.2):

```
PROC GLM;
  CLASS A;                              /A = associate or focal species treatment/
  MODEL X = D A D*A/SS3;                /D = density, use type III sums of squares/
```

Appendix 4.3

Analysis of the ANCOVA model in Eq. (4.3):

```
PROC GLM;
  CLASS A;                          /A = associate or focal species treatment/
  MODEL X = D A/SS3;                          /use type III sums of squares/
```